APPLICATION OF TOPIC MODELING AND DATA VISUALIZATION IN LIBRARIES: A REVIEW OF LITERATURE

Ranjan Borah, *Research Scholar, Department of Library and Information Science, University of Delhi, Delhi-110007*

Email: *ranjanborah87@gmail.com*, **ORCID:** https://orcid.org/0000-0002-0913-2267

Margam Madhusudhan, *Professor, Department of Library and Information Science, University of Delhi, Delhi-110007 Email: madhumargam@gmail.com*
Email: mmadhusudhan@libinfosci.du.ac.in, **ORCID**: https://orcid.org/0009-0004-3412-0526

Abstract

In response to the exponential growth of information, libraries are increasingly adopting advanced computational approaches such as topic modelling, text mining, authorship network analysis, and data visualization. These topics are widely used in different fields and subjects of research, and is constantly evolving its scope and aspects regarding their usability and acceptability, the study has been conducted to clearly understand the present scenario of topic modeling and different techniques of data visualization. These emerging techniques empower users to access precise information while enhancing the efficiency of library services.

This study explores the evolving role of topic modelling and data visualization within Library and Information Science (LIS), focusing on their theoretical underpinnings, practical applications, and recent research trends. To clearly understand the techniques, 73 scholarly publications, published in between 2003-24 were retrieved from different databases such as Scopus, Web of Science, LISA and Google Scholar. Detailed review of the literature of the publications was done to understand the theoretical aspects of the techniques and their future prospects.

Topic modeling is sophisticated text-mining technique that can be done using variety of algorithms such as LDA, STM etc., each one suited to different types of datasets and research objectives, with diverse applications in libraries such as Enhancing Information Retrieval and Search Systems, Improving Digital Library Services, Document Clustering and Recommendation Systems, Subject Metadata and Multimedia Libraries, Topic Trends and Knowledge Discovery in LIS etc.

Analytical findings of the publications reveal that 75% of the literature emerged in the past decade, with a notable concentration between 2018 and 2022, indicating heightened academic interest in these data-driven methodologies. The reviewed documents include journal articles (70%), conference papers (18%), book chapters (9%), and theses (3%), highlighting diverse scholarly contributions.

Topic modelling techniques, particularly Latent Dirichlet Allocation (LDA), are recognized for uncovering hidden patterns in large text corpora, while data visualization translates complex datasets into impactful visuals. Together, these tools offer immense value to LIS professionals, data analysts, and researchers seeking to advance knowledge discovery, improve decision-making, and enhance scholarly communication in modern libraries.

Keywords: Topic Modeling, Latent Dirichlet Algorithm, Data visualization, Library, Librarian.

1.0 Introduction

The insatiable human curiosity for knowledge, coupled with the rapid generation of ideas, development of digital resources, and advancement in research infrastructure, has led to an unprecedented explosion of data. Managing, organizing, and making sense of this vast and complex information, spanning countless disciplines, has become a formidable challenge for libraries and information centres. These institutions are increasingly adopting computational methods powered by algorithms, intelligent systems, and specialized applications to curate, interpret, and analyze large datasets effectively.

The most promising techniques are topic modelling, text mining, journal mapping, authorship network analysis, and data visualization. These techniques evolved during the last two decades and a lot of research revolving around these topics is constantly being carried out at different levels and domains. Since these techniques are widely used in different fields and subjects of research, and is constantly evolving its scope and aspects regarding their usability and acceptability, the study has been conducted to clearly understand the present scenario of topic modeling and different techniques of data visualization. Library and Information Science is one such domain where new ideas are constantly incorporated and tested for the smooth functioning, and benefits of the information centres to efficiently work for their users. Analyzing the metadata of relevant scholarly documents aims to provide a comprehensive literature review and demonstrate how topic modelling can be effectively leveraged to enhance library services and information discovery. It will help to understand how these innovative tools can enable libraries to deliver more precise, user-centred information services and help them to transform raw data into actionable insights.

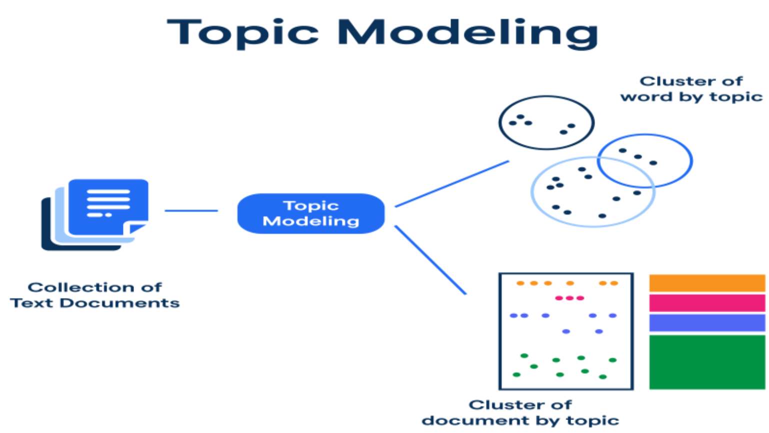

Figure 1: Image depicting the process of Topic Modeling

2.0 Methodology

The undertaken study is conducted to clearly understand the various hidden concepts related with topic modeling, the algorithms used for conducting topic modeling, how to use visualization techniques to visualize the relevant information, and explore the application and relevance of topic modelling and data visualization within the library and information science domain.

Since the study is related to topic modeling and use of visualisation of data, searches were carried out across multiple academic databases and scholarly platforms, including Library and Information Science Abstracts (LISA), Scopus, Web of Science, Emerald Insight, and Google Scholar**,** in addition to ResearchGate and publisher-specific e-journal websites by using targeted keyword combinations such as *"topic model*", *"LDA" (Latent Dirichlet Allocation)*, *"STM" (Structural Topic Model)*, *"data visualization"*, and *"topic modelling in libraries"* to retrieve relevant literature. 89 documents were identified as potentially relevant for the study from the initial pool of results of 200+ documents. After a detailed screening for quality, scope, and relevance to the research objectives, 73 documents published between 2003 and 2024 were selected for in-depth analysis.

The detailed review of literature of the documents of topic modeling and data visualization was done separately to understand the key concepts clearly. Their potential use and future prospects in libraries were also studied. The selected corpus was examined using ‘R’ programming language within the RStudio environment to conduct bibliometric and content-based analyses. The theoretical approach enabled qualitative understanding of the relevant topics and trends, and the analytical approach enabled quantitative and thematic visualization of the field's trends, authorship patterns, and topic evolution.

3.0 Topic Modeling (TM)

Topic modelling is a powerful text-mining technique to uncover hidden thematic structures within large collections of unstructured documents. It works by analyzing patterns in word usage and grouping frequently co-occurring terms into distinct topics, thus enabling researchers to identify dominant themes across vast datasets. This technique facilitates the organization and interpretation of complex textual information and enhances the generation of meaningful insights.

At its core, topic modelling relies on probabilistic models that process textual data to extract coherent topics, offering a structured representation of unstructured content. As Blei (2012) explains, *"Topic models are algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents and can organize the collection according to the discovered themes. Topic modelling algorithms can be applied to massive collections of documents and can be adapted to many kinds of data. They can find patterns in genetic data, images, and social networks, among other applications."*

In the research context, topic modelling is a valuable tool for tracking evolving trends, identifying emerging areas, and discovering new fields of inquiry. As Lamba and Madhusudhan (2019) noted, *"Topic modelling not only helps the researcher to determine the trending themes or related fields concerning their field of interest but also helps them to identify new concepts and fields over time."*

While topic modelling has recently gained traction in Library and Information Science (LIS), it has long been employed in computer science, digital humanities, biomedical engineering, and medicine. Its diverse applications include document classification, literature reviews, citation analysis, keyword extraction, recommendation systems, marketing analysis, and similarity searches.

A variety of sophisticated algorithms are available for implementing topic modelling. Among the most widely used are Latent Dirichlet Allocation (LDA), Structural Topic Modeling (STM), Dynamic Topic Modeling, Wikipedia-based Polyglot Dirichlet Allocation (WikiPDA), Latent Semantic Analysis (LSA), and BERTopic. Each algorithm offers unique strengths suited to different types of datasets and research objectives, making topic modelling a versatile and indispensable method in modern data analysis.

3.1 Review of Related Literature

Blei et al. (2003) laid the foundation of topic modelling and introduced the “Latent Dirichlet Allocation (LDA) model, an influential generative probabilistic framework for discrete data such as text corpora. They described LDA as a three-level hierarchical Bayesian model, where each document is treated as a mixture of topics and each topic as a distribution over words.” This model became the cornerstone for subsequent topic modelling and text-mining developments.

Expanding on this, Wallach (2006) proposed a hierarchical model integrating n-gram statistics with latent topic variables, thereby improving the interpretability of inferred topics. Using the Gibbs EM algorithm on real datasets, Wallach demonstrated that this model outperformed unigram-based approaches and produced more semantically meaningful results, especially by minimizing the dominance of function words.

Nallapati et al. (2008) extended topic modelling to include citation analysis through two novel models: Pairwise-Link-LDA and Link-PLSALDA. These models allowed joint modelling of text and citation networks, enhancing the representation of scholarly communication, albeit with trade-offs in computational complexity.

In the same year, Titov and McDonald (2008) introduced a multi-grain topic modelling framework to extract ratable aspects from online reviews. Their method addressed the limitations of traditional models by distinguishing between global object properties and fine-grained aspects typically evaluated by users.

Bridging digital libraries with computational linguistics, Mehler and Waltinger (2009) used Dewey Decimal Classification (DDC) and metadata from the Open Archives Initiative (OAI) to create a classification model. Their work demonstrated the potential of topic models for automated thematic categorization in digital libraries.

Hu et al. (2014) identified the gap between machine learning-driven topic models and their practical usability by non-experts. They introduced Interactive Topic Modeling (ITM), a user-centric framework that encodes feedback into topic models, thus making them more adaptable and interpretable.

That same year, Roberts et al. (2014) introduced the Structural Topic Model (STM), which integrates document metadata into the modelling process. This semi-automated technique was especially useful for analyzing open-ended survey responses, estimating treatment effects, and understanding respondent behaviour.

Toraman and Can (2015) developed a priority-based algorithm for selecting news articles based on their importance and diversity using topic modelling. This was aimed at optimizing public news front pages in the absence of meta-attributes like click counts.

To address social data's multi-modal nature, Qian et al. (2016) proposed the Multi-modal Event Topic Model (mmETM) to capture both textual and visual elements from social media. Their framework effectively identified topic evolution and event summaries, outperforming several state-of-the-art techniques.

Nikolenko et al. (2017) tackled two key limitations in topic modelling: quality assessment and sub-topic detection. They proposed tf-idf coherence for more accurate topic evaluation. They introduced the Interval Semi-Supervised LDA (ISLDA) for focused sub-topic mining, validating their approach with a case study on ethnicity discourse in Russian blogs.

Agrawal et al. (2018) exposed a critical flaw in LDA, order effects, whereby reshuffling input data leads to inconsistent topic outputs. This discovery raised concerns about reproducibility and accuracy in topic-based studies.

A methodological framework for rigorous LDA implementation in communication research was developed by Maier et al. (2018). Their study emphasized reliability and validity, showcasing LDA as a robust tool combined with appropriate preprocessing and evaluation strategies.

Providing a panoramic overview, Kherwa and Bansal (2018) conducted a comprehensive survey of approximately 300 articles on topic modelling. They classified topic modelling techniques, posterior inference methods, LDA variants, and applications across domains such as bioinformatics, social networks, and scientific literature.

Hagen (2018) evaluated the utility of LDA in analyzing e-petitions, demonstrating that over 87% of topics generated were interpretable to human judges. He highlighted the advantages of LDA over manual content analysis, such as bias reduction and discovery of latent themes.

Hannigan et al. (2019) advocated for topic modelling in management scholarship, noting its capacity to uncover novelty, support inductive classification, and understand cultural dynamics. They positioned topic modelling as a bridge between algorithmic analysis and qualitative insight.

Vayansky and Kumar (2020) reviewed emerging topic modelling approaches capable of handling short, sparse texts, temporal topic evolution, and topic correlations. They proposed a decision tree to help researchers choose suitable models based on dataset characteristics and analytical goals.

Dieng et al. (2020) introduced the Embedded Topic Model (ETM) to address the limitations of traditional models with large and skewed vocabularies. Using amortized variational inference, ETM delivered interpretable and high-performing results even with rare and high-frequency words.

Qiao and Williams (2021) applied LDA and sentiment analysis to explore discussions on global warming on Twitter. Their study revealed dominant thematic clusters and a higher prevalence of positive sentiment in online discourse.

In a comparative study, Bellaouar et al. (2021) evaluated the performance of LDA and Latent Semantic Analysis (LSA) using NIPS conference papers, analyzing the impact of text preprocessing techniques like lemmatization and bi-gram collocation.

Ballester and Penner (2022) proposed new metrics to assess topic models' robustness, descriptive power, and realism. They found that neural paragraph embedding methods could maintain stable and scalable document-topic representations beyond the capabilities of traditional models.

Qiang et al. (2022) systematically reviewed short-text topic modelling, categorizing approaches into three types: Dirichlet multinomial mixture-based, global co-occurrence-based, and self-aggregation-based. Their work also led to the development of STTM, the first open-source Java library for short-text topic modelling.

Ravenda et al. (2022) employed topic modelling to analyze municipal communication on Facebook, uncovering themes in social media posts from 502 Italian municipalities between 2016 and 2018.

In systematic reviews, Natukunda and Muchene (2023) developed a statistical model using LDA for automated title and abstract screening, enhancing the speed and accuracy of evidence synthesis processes.

Finally, Mamasodikova and Ergasheva (2024) applied topic modelling to explore gender disparities in news coverage using an ever-expanding open news corpus. Their visualization-enabled methodology successfully tracked topic evolution and key real-world events over time using monthly generated LDA topics.

3.2 Application of Topic Modeling in Libraries

Topic modelling has increasingly proven valuable for organizing, analyzing, and enhancing access to information in libraries and digital collections. Its applications span from improving search and retrieval functions to identifying research trends and user needs.

(i) **Enhancing Information Retrieval and Search Systems:** Muresan and Harper (2004) demonstrated how topic modelling could generate "mediated queries" that more accurately represent users' information needs, enabling efficient search across large, diverse collections. Similarly, Haruechaiyasak and Damrongrat (2008) used LDA on Wikipedia articles to recommend highly relevant content based on topic similarity, even beyond hyperlink connections.

(ii) **Improving Digital Library Services:** Bethard et al. (2009) combined topic modelling with classification techniques to select resources for subject-based digital libraries. Newman et al. (2010) highlighted how topic models can automate subject tagging and improve content discovery in digital collections. Efron et al. (2011) addressed the challenge of metadata inconsistency in large collections and proposed methods to identify low-information documents to refine topic modelling outputs.

(iii) **Document Clustering and Recommendation Systems:** Wang and Jie (2012) applied probabilistic topic modelling with Gibbs sampling for accurate clustering of digital books. Scherer et al. (2013) introduced a topic-based similarity function for efficient indexing in public library collections. Charlin et al. (2014) proposed the Collaborative Score Topic Model (CSTM), which combines user preferences and textual content for personalized document recommendations.

(iv) **Subject Metadata and Multimedia Libraries:** Choi et al. (2015) used LDA to analyze song interpretations, showing how topic modelling can enrich metadata in music libraries and potentially for literary collections. Cain (2016) emphasized how topic modelling can improve access to poorly indexed digital texts by identifying latent content structures.

(v) **Domain-Specific Topic Exploration and Visualization:** Mesbah et al. (2017) introduced "Facet Embedding," a domain-aware model for generating metadata and visualizing scientific trends. Abuhay et al. (2017) analyzed 16 years of conference papers to uncover topic evolution and interdisciplinary links through dynamic topic networks. Figuerola et al. (2017) applied LDA to the LISA database to identify core LIS research areas over nearly four decades.

(vi) **Topic Trends and Knowledge Discovery in LIS:** Kurata et al. (2018) used LDA to uncover hidden thematic structures within LIS literature. Al-Natsheh et al. (2018) addressed the issue of community-specific keywords and proposed a semantic tagging system using topic modelling for improved retrieval. Fang et al. (2018) combined topic modelling with regression analysis to identify trending and declining topics in LIS.

(vii) **User Behavior and Service Optimization:** Chen and Wang (2019) analyzed library chat transcripts using LDA and discovered that access to library resources was a dominant concern among users. Joo et al. (2020) studied Facebook posts from public libraries and found that posts about events and multimedia content had higher user engagement, using topic modelling to classify content themes.

(viii) **Evaluation, Research Trends, and Visualization:** Papachristopoulos et al. (2019) integrated topic modelling with citation and altmetrics analysis to map the structure of digital library evaluation literature. Viet and Kravets (2019) scraped and analyzed arXiv papers to detect emerging academic trends. Timakum et al. (2020) applied text mining and topic modelling on top LIS journals to map evolving research areas. Meanwhile, Miyata et al. (2020) found a shift in LIS research focus toward digital and internet-based topics post-2015.

(ix) **Language and Cross-Lingual Applications:** Piccardi and West (2021) introduced the Wikipedia-based Polyglot Dirichlet Allocation (WikiPDA), enabling cross-lingual topic modelling for multilingual content classification.

(x) **Classification and Indexing of Unstructured Texts:** Mohammed and Al-Augby (2020) compared LDA and LSA for classifying 23 million words from unstructured scientific texts, concluding LDA offered superior coherence and clustering accuracy. Mazumder and Barui (2021) analyzed titles of LIS theses from Shodhganga using LDA to extract dominant research themes.

(xi) **Recommendation Systems and Conference Analysis:** Jelodar et al. (2021) applied topic modelling to eight computer science conference publications to build a recommendation system aligned with future research directions. Dantu et al. (2021) used topic modelling with co-citation analysis to explore research on healthcare applications of IoT, revealing five key thematic clusters.

(xii) **LIS Theses and Doctoral Research Analysis:** Kumar and Thakur (2022) employed topic modelling on Indian LIS doctoral theses to categorize them into ten major research domains, highlighting ICT as a key area.

(xiii) **Topic Modeling for Visualizing and Comparing Research:** Saha and Ghosh (2023) analyzed 734 LIS articles from Indian and international journals using topic modelling and visualization, identifying distinct research priorities between the two.

(xiv) **Librarian Decision-Making and Reference Analysis:** Koh and Fienup (2021) evaluated different topic modelling algorithms for chat reference data and found that probabilistic latent semantic analysis (LSA) outperformed LDA in interpretability and coherence.

(xv) **Authorship, Verification, and Textual Analysis:** Potha and Stamatatos (2019) explored topic modelling in authorship verification. They found that LDA was more effective when combined with profile-based and extrinsic verification methods.

(xvi) **Indian Contextual Studies:** Lamba and Madhusudhan (2019) examined nearly four decades of Indian LIS research, identifying bibliometrics, ICT, and user studies as dominant themes.

4.0 Data Visualization (DV)

Data visualization (DV) refers to transforming abstract, non-coordinate data into meaningful graphical or pictorial formats. It allows users to interpret complex datasets more efficiently than traditional text-based methods. As Lamba and Madhusudhan (2022) note, "Visualization is a method of externalizing information, making it visible, interactive, and easier to manipulate through visual systems." This aligns with the well-known adage, *"A picture is worth a thousand words"*, emphasizing how visuals communicate data-driven narratives more effectively than dense text (Finch & Flenner, 2016).

Zakaria (2021) highlights that graphical representations are cognitively more persuasive than numerical tables, enabling faster comprehension of data relationships. Embarak (2018) further explains that DV simplifies massive datasets into digestible visuals, aiding clearer communication and insight generation. Similarly, Pant and Rajput (n.d.) stress that DV is a powerful means of describing large-scale data in accessible and engaging ways.

Sadiku et al. (2016) recognize DV as a universally adaptable tool with speed, simplicity, and cross-disciplinary applicability. According to Lamba and Madhusudhan (2022), DV also supports problem-solving, data interpretation, and insight sharing—capabilities enhanced through interactive and computational tools.

4.1 Data Visualization in Libraries

In library and information science, data visualization is rapidly emerging as an essential practice for managing and interpreting complex information resources. Siddiqui (2021) emphasizes that DV provides intuitive visual formats, charts, graphs, and maps that reveal trends and patterns across large datasets, making data more accessible to library users.

Midway (2020) reminds us, however, that the art of data visualization must go hand-in-hand with accuracy. Misrepresentations, even if unintentional, can mislead users and obscure insights. Hence, thoughtful design and precision are vital.

Murphy (2015) notes that the growing complexity of digital content and aggregated data has made DV a necessary tool in academic libraries. It is now employed for collection analysis, usage tracking, expenditure visualization, and overall service enhancement. Finch and Flenner (2016) add that DV can support decision-making by illuminating citation patterns, suggesting titles for weeding, mapping scholarly communication, and analyzing subject relationships.

As Aavarti and Kumbar (2022) state, DV tools offer academic libraries new opportunities to contribute to research by integrating visualization into regular services. These tools can enhance library value by presenting data in ways that engage faculty, researchers, and students alike. Eaton (2017) further asserts that academic libraries are well-positioned to leverage ICT and DV technologies to deliver impactful research support.

Murphy (2013) points out that DV strengthens the library's role in research support, enabling librarians to organize, interpret, and present data meaningfully. Aavarti and Kumbar (2022) also recognize that while academic libraries are still developing their capabilities in this domain, DV holds immense potential in visualizing research outputs, enabling deeper analytics, and supporting user needs through charts, infographics, networks, and more.

According to Aghassibake et al. (2020), offering DV instruction also presents a chance to address broader pedagogical goals, particularly emerging digital tools and evolving librarian roles. Mahajan and Gokhale (2020) classify the advantages of DV into two broad areas: (i) visual impact, making data memorable and compelling, and (ii) functionality, improving data navigation, decision-making, and service planning.

Murphy (2013) and Brigham (2016) emphasize that DV empowers librarians to manage and present complex data from varied sources creatively and strategically. It assists with strategic planning, reveals user behaviour, supports evidence-based decisions, and visualizes institutional priorities.

Eaton (2017) adds that DV enables librarians to gain deeper insights from data, thereby enhancing their contributions to research and institutional goals. Zakaria (2021) further notes that DV demonstrates the library's value and markets librarians' expertise effectively. Aavarti and Kumbar (2022) affirm that globally, academic libraries have embraced DV support services, offering consultancy, workshops, training, and dedicated DV spaces to foster research and innovation.

5.0 Findings of the Study

The key findings from the analysis of the publications are presented below, supplemented with tables and graphical visualizations to highlight trends and patterns.

(i) **Publication Timeline (2003–2024):** A total of 73 scholarly works spanning two decades were analyzed. Notably, 75% of the publications were released within the last ten years, reflecting a sharp increase in interest and application of topic modelling and data visualization in recent times. The majority of these contributions were concentrated between 2018 and 2022, indicating the subject area's current and growing relevance. (Figure 2).

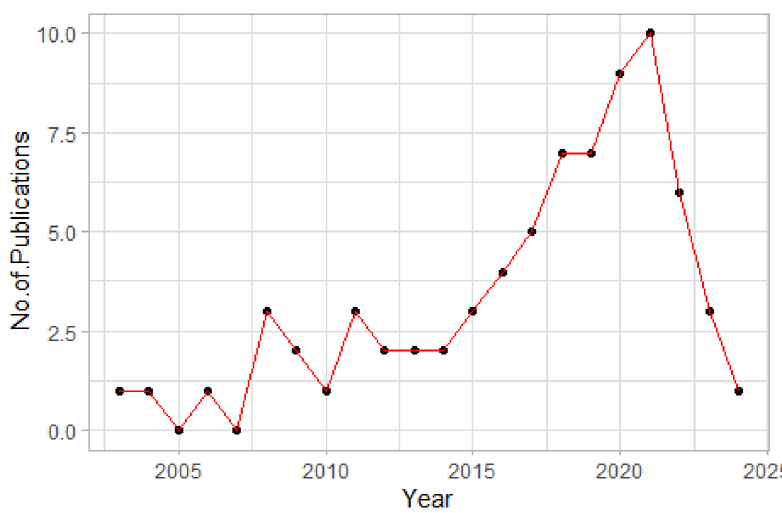

Figure 2: Year-wise distribution of publications

(ii) **Document Types:** The review encompassed four major categories of publications: journal articles, conference proceedings, book chapters, and theses.

- **Research articles** constituted the largest share (70%), demonstrating the prominence of peer-reviewed scholarly communication in this field.
- **Conference papers** accounted for 18%, followed by book chapters (9%) and doctoral theses (3%), indicating a growing academic interest across diverse platforms (Table 1).

Table 1: Types of publications

Types of publications	No. of Publications
Journal Articles	51
Book chapters	7
Conference Papers	13
Theses	2

(iii) **Authorship Patterns:** The documents analyzed exhibited various authorship patterns, ranging from single-author contributions to large collaborative works.

- Publications with two authors were the most common (35%, or 26 papers).

- Single-author and three-author publications accounted for 18% each (13 papers each).
- A smaller number of documents were found to have 7, 8, and even 11 authors**,** reflecting interdisciplinary and collaborative research practices in some cases.
- (Figure 3).

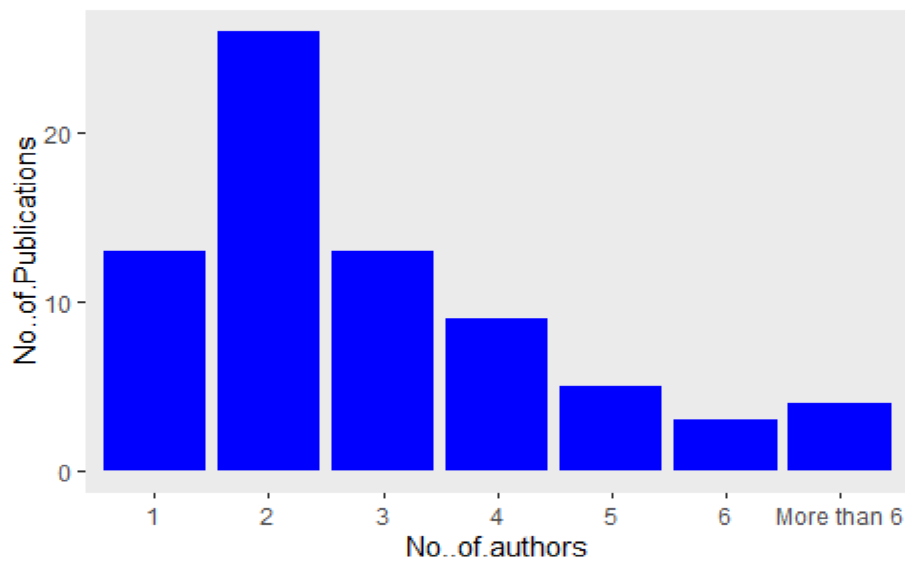

Figure 3: Authorship Pattern of Publications

These findings reflect the dynamic and interdisciplinary nature of topic modelling and data visualization research in the library science context, with a clear trend towards a collaborative, data-intensive scholarship in recent years.

7.0 Conclusion

Topic modelling is an emerging and versatile technique that, when applied effectively, can benefit librarians and other information professionals. This study explores theoretical and practical aspects of topic modelling within the Library and Information Science (LIS) field, including various algorithms suitable for analyzing datasets. It examines the advantages, limitations, uses, challenges, and real-world applications of topic modelling. It helps the libraries in various day to day applications ranging from creation of document clustering and recommendation systems, subject metadata and multimedia libraries, to discovering trending topics and knowledge in LIS, studying the user behaviour and service optimization, for cross-lingual applications, classification and indexing of unstructured texts, and can be used for analysis of LIS theses, doctoral research, conference etc. Additionally, the study highlights the importance of data visualization in presenting complex and large datasets in clear, accurate, and efficient graphical formats. Visual representations are particularly impactful, as people tend to retain and prioritize visual information over raw data. As such, data visualization is widely valued across academic disciplines. This study aims to support librarians, data analysts, and researchers in understanding and applying topic modelling and visualization techniques to enhance library services, contribute to scholarly communication, and foster their professional growth.

References

Aavarti, S., & Kumbar, R. (2022). *Data Visualization Software Tools for Academic Libraries: An Analysis*.

Abuhay, T. M., Kovalchuk, S. V., Bochenina, K. O., Kampis, G., Krzhizhanovskaya, V. V., & Lees, M. H. (2017). Analysis of Computational Science Papers from ICCS 2001-2016 using Topic Modeling and Graph Theory. *Procedia Computer Science*, *108*, 7–17. https://doi.org/10.1016/j.procs.2017.05.183

Aghassibake, N., Joque, J., & Sisk, M. L. (2020). *Supporting Data Visualization Services in Academic Libraries*. *18*. https://jitp.commons.gc.cuny.edu/supporting-data-visualization-services-in-academic-libraries/

Agrawal, A., Fu, W., & Menzies, T. (2018). What is wrong with topic modelling? And how to fix it using search-based software engineering. *Information and Software Technology*, *98*, 74–88. https://doi.org/10.1016/j.infsof.2018.02.005

Al-Natsheh, H. T., Martinet, L., Muhlenbach, F., Rico, F., & Zighed, D. A. (2018). Metadata Enrichment of Multi-disciplinary Digital Library: A Semantic-Based Approach. In E. Méndez, F. Crestani, C. Ribeiro, G. David, & J. C. Lopes (Eds.), *Digital Libraries for Open Knowledge* (Vol. 11057, pp. 32–43). Springer International Publishing. https://doi.org/10.1007/978-3-030-00066-0_3

Ballester, O., & Penner, O. (2022). Robustness, replicability and scalability in topic modeling. *Journal of Informetrics*, *16*(1), 101224. https://doi.org/10.1016/j.joi.2021.101224

Bellaouar, S., Bellaouar, M. M., & Ghada, I. E. (2021). Topic Modeling: Comparison of LSA and LDA on Scientific Publications. *2021 4th International Conference on Data Storage and Data Engineering*, 59–64. https://doi.org/10.1145/3456146.3456156

Bethard, S., Ghosh, S., Martin, J. H., & Sumner, T. (2009). Topic model methods for automatically identifying out-of-scope resources. *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 19–28. https://doi.org/10.1145/1555400.1555405

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*.

Brigham, T. J. (2016). Feast for the Eyes: An Introduction to Data Visualization. *Medical Reference Services Quarterly*, *35*(2), 215–223. https://doi.org/10.1080/02763869.2016.1152146

Cain, J. O. (2016). Using Topic Modeling to Enhance Access to Library Digital Collections. *Journal of Web Librarianship*, *10*(3), 210–225. https://doi.org/10.1080/19322909.2016.1193455

Charlin, L., Zemel, R. S., & Larochelle, H. (2014). Leveraging user libraries to bootstrap collaborative filtering. *Proceedings of the 20th ACM SIGKDD International Conference* *on Knowledge Discovery and Data Mining*, 173–182. https://doi.org/10.1145/2623330.2623663

Chen, X., & Wang, H. (2019). Automated chat transcript analysis using topic modeling for library reference services. *Proceedings of the Association for Information Science and Technology*, *56*(1), 368–371. https://doi.org/10.1002/pra2.31

Choi, K., Lee, J. H., Willis, C., & Downie, J. S. (2015). Topic Modeling Users' Interpretations of Songs to Inform Subject Access in Music Digital Libraries. *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, 183–186. https://doi.org/10.1145/2756406.2756936

Dantu, R., Dissanayake, I., & Nerur, S. (2021). Exploratory Analysis of Internet of Things (IoT) in Healthcare: A Topic Modeling & Co-citation Approaches. *Information Systems Management*, *38*(1), 62–78. https://doi.org/10.1080/10580530.2020.1746982

Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2020). Topic Modeling in Embedding Spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439–453. https://doi.org/10.1162/tacl_a_00325

Eaton, M. (2017). Seeing Library Data: A Prototype Data Visualization Application for Librarians. *Journal of Web Librarianship*, *11*(1), 69–78. https://doi.org/10.1080/19322909.2016.1239236

Efron, M., Organisciak, P., & Fenlon, K. (2011). Building topic models in a federated digital library through selective document exclusion. *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1–10. https://doi.org/10.1002/meet.2011.14504801048

Embarak, O. (2018). The Importance of Data Visualization in Business Intelligence. In Dr. O. Embarak (Ed.), *Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems* (pp. 85–124). Apress. https://doi.org/10.1007/978-1-4842-4109-7_2

Fang, D., Yang, H., Gao, B., & Li, X. (2018). Discovering research topics from library electronic references using latent Dirichlet allocation. *Library Hi Tech*, *36*(3), 400–410. https://doi.org/10.1108/LHT-06-2017-0132

Figuerola, C. G., García Marco, F. J., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. *Scientometrics*, *112*(3), 1507–1535. https://doi.org/10.1007/s11192-017-2432-9

Finch, J. L., & Flenner, A. R. (2016). Using Data Visualization to Examine an Academic Library Collection. *College & Research Libraries*, *77*(6), 765–778. https://doi.org/10.5860/crl.77.6.765

Hagen, L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing & Management*, *54*(6), 1292–1307. https://doi.org/10.1016/j.ipm.2018.05.006

Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic Modeling in Management Research: Rendering New Theory from Textual Data. *Academy of Management Annals*, *13*(2), 586–632. https://doi.org/10.5465/annals.2017.0099

Haruechaiyasak, C., & Damrongrat, C. (2008). Article Recommendation Based on a Topic Model for Wikipedia Selection for Schools. In G. Buchanan, M. Masoodian, & S. J. Cunningham (Eds.), *Digital Libraries: Universal and Ubiquitous Access to Information* (Vol. 5362, pp. 339–342). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-89533-6_39

Hu, Y., Boyd-Graber, J., Satinoff, B., & Smith, A. (2014). Interactive topic modeling. *Machine Learning*, *95*(3), 423–469. https://doi.org/10.1007/s10994-013-5413-0

Jelodar, H., Wang, Y., Xiao, G., Rabbani, M., Zhao, R., Ayobi, S., Hu, P., & Masood, I. (2021). Recommendation system based on semantic scholar mining and topic modeling on conference publications. *Soft Computing*, *25*(5), 3675–3696. https://doi.org/10.1007/s00500-020-05397-3

Joo, S., Lu, K., & Lee, T. (2020). Analysis of content topics, user engagement and library factors in public library social media based on text mining. *Online Information Review*, *44*(1), 258–277. https://doi.org/10.1108/OIR-11-2018-0345

Kherwa, P., & Bansal, P. (2018). Topic Modeling: A Comprehensive Review. *ICST Transactions on Scalable Information Systems*, *0*(0), 159623. https://doi.org/10.4108/eai.13-7-2018.159623

Koh, H., & Fienup, M. (2021). Topic Modeling as a Tool for Analyzing Library Chat Transcripts. *Information Technology and Libraries*, *40*(3). https://doi.org/10.6017/ital.v40i3.13333

Kumar, V., & Thakur, K. (2022). Using text analysis to study doctoral-level library and information science research trends in India. *Annals of Library and Information Studies*, *69*(3). https://doi.org/10.56042/alis.v69i3.58719

Kumari, R., Jeong, J. Y., Lee, B.-H., Choi, K.-N., & Choi, K. (2021). Topic modeling and social network analysis of publications and patents in humanoid robot technology. *Journal of Information Science*, *47*(5), 658–676. https://doi.org/10.1177/0165551519887878

Kurata, K., Miyata, Y., Ishita, E., Yamamoto, M., Yang, F., & Iwase, A. (2018). Analyzing library and information science full-text articles using a topic modeling approach. *Proceedings of the Association for Information Science and Technology*, *55*(1), 847–848. https://doi.org/10.1002/pra2.2018.14505501143

Lamba, M., & Madhusudhan, M. (2019). Mapping of topics in DESIDOC Journal of Library and Information Technology, India: A study. *Scientometrics*, *120*(2), 477–505. https://doi.org/10.1007/s11192-019-03137-5

Lamba, M., & Madhusudhan, M. (2022). Information Visualization. In M. Lamba & M. Madhusudhan (Eds.), *Text Mining for Information Professionals: An Uncharted Territory* (pp. 243–293). Springer International Publishing. https://doi.org/10.1007/978-3-030-85085-2_9

Mahajan, K. N., & Gokhale, L. A. (2020). *Comparative Study of Data Visualization Tools*. 48–53. https://www.researchgate.net/publication/344425307_Comparative_Study_of_Data_Visualization_Tools

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Mazumder, S., & Barui, T. (2021). *Discovering Topics from the Titles of the Indian LIS Theses*.

Mehler, A., & Waltinger, U. (2009). Enhancing document modeling by means of open topic models: Crossing the frontier of classification schemes in digital libraries by example of the DDC. *Library Hi Tech*, *27*(4), 520–539. https://doi.org/10.1108/07378830911007646

Mesbah, S., Fragkeskos, K., Lofi, C., Bozzon, A., & Houben, G.-J. (2017). Facet Embeddings for Explorative Analytics in Digital Libraries. In J. Kamps, G. Tsakonas, Y. Manolopoulos, L. Iliadis, & I. Karydis (Eds.), *Research and Advanced Technology for Digital Libraries* (Vol. 10450, pp. 86–99). Springer International Publishing. https://doi.org/10.1007/978-3-319-67008-9_8

Midway, S. R. (2020). Principles of Effective Data Visualization. *Patterns*, *1*(9), 1–7. https://doi.org/10.1016/j.patter.2020.100141

Miyata, Y., Ishita, E., Yang, F., Yamamoto, M., Iwase, A., & Kurata, K. (2020). Knowledge structure transition in library and information science: Topic modeling and visualization. *Scientometrics*, *125*(1), 665–687. https://doi.org/10.1007/s11192-020-03657-5

Mohammed, S. H., & Al-augby, S. (2020). LSA & LDA topic modeling classification: Comparison study on e-books. *Indonesian Journal of Electrical Engineering and Computer Science*, *19*(1), 353. https://doi.org/10.11591/ijeecs.v19.i1.pp353-362

Muresan, G., & Harper, D. J. (2004). Topic modeling for mediated access to very large document collections. *Journal of the American Society for Information Science and Technology*, *55*(10), 892–910. https://doi.org/10.1002/asi.20034

Murphy, S. A. (2013). Data Visualization and Rapid Analytics: Applying Tableau Desktop to Support Library Decision-Making. *Journal of Web Librarianship*, *7*(4), 465–476. https://doi.org/10.1080/19322909.2013.825148

Murphy, S. A. (2015). How data visualization supports academic library assessment: Three examples from The Ohio State University Libraries using Tableau. *College & Research Libraries News*, *76*(9), 482–486. https://doi.org/10.5860/crln.76.9.9379

Nallapati, R. M., Ahmed, A., Xing, E. P., & Cohen, W. W. (2008). Joint latent topic models for text and citations. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 542–550. https://doi.org/10.1145/1401890.1401957

Natukunda, A., & Muchene, L. K. (2023). Unsupervised title and abstract screening for systematic review: A retrospective case-study using topic modeling methodology. *Systematic Reviews*, *12*(1), 1. https://doi.org/10.1186/s13643-022-02163-4

Newman, D., Noh, Y., Talley, E., Karimi, S., & Baldwin, T. (2010). Evaluating topic models for digital libraries. *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, 215–224. https://doi.org/10.1145/1816123.1816156

Nikolenko, S. I., Koltcov, S., & Koltsova, O. (2017). Topic modeling for qualitative studies. *Journal of Information Science*, *43*(1), 88–102. https://doi.org/10.1177/0165551515617393

Pant, A., & Rajput, R. S. (n.d.). *Introduction To Research Data And Its Visualization Using R*.

Papachristopoulos, L., Tsakonas, G., Boudourides, M., Sfakakis, M., Kleidis, N., Lenis, S., & Papatheodorou, C. (2019). Discovering the structure and impact of the digital library evaluation domain. *International Journal on Digital Libraries*, *20*(2), 125–141. https://doi.org/10.1007/s00799-017-0222-x

Piccardi, T., & West, R. (2021). Crosslingual Topic Modeling with WikiPDA. *Proceedings of the Web Conference 2021*, 3032–3041. https://doi.org/10.1145/3442381.3449805

Potha, N., & Stamatatos, E. (2019). Improving author verification based on topic modeling. *Journal of the Association for Information Science and Technology*, *70*(10), 1074–1088. https://doi.org/10.1002/asi.24183

Qian, S., Zhang, T., Xu, C., & Shao, J. (2016). Multi-Modal Event Topic Model for Social Event Analysis. *IEEE Transactions on Multimedia*, *18*(2), 233–246. https://doi.org/10.1109/TMM.2015.2510329

Qiang, J., Qian, Z., Li, Y., Yuan, Y., & Wu, X. (2022). Short Text Topic Modeling Techniques, Applications, and Performance: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, *34*(3), 1427–1445. https://doi.org/10.1109/TKDE.2020.2992485

Qiao, F., & Williams, J. (2021). Topic Modeling and Sentiment Analysis of Global Warming Tweets: Evidence From Big Data Analysis. *Journal of Organizational and End User Computing*, *34*(3), 1–18. https://doi.org/10.4018/JOEUC.294901

Ravenda, D., Valencia-Silva, M. M., Argiles-Bosch, J. M., & García-Blandón, J. (2022). The strategic usage of Facebook by local governments: A structural topic modeling analysis. *Information & Management*, *59*(8), 103704. https://doi.org/10.1016/j.im.2022.103704

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Sadiku, M., Shadare, A., Musa, S., Akujuobi, C., & Perry, R. (2016). Data visualization. *International Journal of Engineering Research and Advanced Technology (IJERAT)*, *12*, 2454–6135.

Saha, M., & Ghosh, S. (2023). Topic Modeling in Library and Information Science from the Primary Data: Swing in Thrust Areas. *International Journal of Information Science and Management (IJISM)*, *21*(3). https://doi.org/10.22034/ijism.2023.1977569.0

Scherer, M., Von Landesberger, T., & Schreck, T. (2013). Topic Modeling for Search and Exploration in Multivariate Research Data Repositories. In T. Aalberg, C. Papatheodorou, M. Dobreva, G. Tsakonas, & C. J. Farrugia (Eds.), *Research and Advanced Technology for Digital Libraries* (Vol. 8092, pp. 370–373). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40501-3_39

Shokhistakhon Mamasodikova & Guli Ergasheva. (2024). *Gender bias in the news: topic modeling and visualization framework*. https://doi.org/10.13140/RG.2.2.11765.35044

Siddiqui, A. T. (2021). Data Visualization: A Study of Tools and Challenges. *Asian Journal of Technology & Management Research*, *11*(1), 18–23. https://doi.org/10.1016/j.patter.2020.100141

Timakum, T., Kim, G., & Song, M. (2020). A data-driven analysis of the knowledge structure of library science with full-text journal articles. *Journal of Librarianship and Information Science*, *52*(2), 345–365. https://doi.org/10.1177/0961000618793977

Titov, I., & McDonald, R. (2008). Modeling online reviews with multi-grain topic models. *Proceedings of the 17th International Conference on World Wide Web*, 111–120. https://doi.org/10.1145/1367497.1367513

Toraman, C., & Can, F. (2015). A front-page news-selection algorithm based on topic modeling using raw text. *Journal of Information Science*, *41*(5), 676–685. https://doi.org/10.1177/0165551515589069

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. https://doi.org/10.1016/j.is.2020.101582

Viet, N. T., & Kravets, A. G. (2019). Analyzing Recent Research Trends of Computer Science from Academic Open-access Digital Library. *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, 31–36. https://doi.org/10.1109/SMART46866.2019.9117215

Wallach, H. M. (2006). Topic modeling: Beyond bag-of-words. *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, 977–984. https://doi.org/10.1145/1143844.1143967

Wang, L., & Jie, Y. (2012). Enhancing Digital Book Clustering by LDAC Model. *IEICE Transactions on Information and Systems*, *E95.D*(4), 982–988. https://doi.org/10.1587/transinf.E95.D.982

Zakaria, M. S. (2021). Data visualization as a research support service in academic libraries: An investigation of world-class universities. *The Journal of Academic Librarianship*, *47*(5), 102397. https://doi.org/10.1016/j.acalib.2021.102397